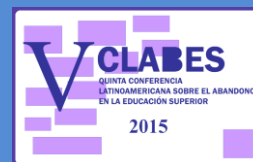




V CLABES
QUINTA CONFERENCIA
LATINOAMERICANA SOBRE EL
ABANDONO EN LA EDUCACIÓN
SUPERIOR



UN MODELO PARA EXPLICAR LA RETENCIÓN EN LA UNIVERSIDAD DE
BOGOTÁ JORGE TADEO LOZANO: ARBOLES DE DECISIÓN.

Línea 1: Factores asociados al abandono. Tipos y perfiles de abandono: Modelos y/o métodos de medición de las causas asociadas al abandono.

BARRAGÁN, Sandra¹

GONZÁLEZ, Leandro²

Universidad de Bogotá Jorge Tadeo Lozano - Colombia

e-mail: sandra.barragan@utadeo.edu.co

Resumen. En el seguimiento a la no matrícula y a la retención estudiantil en la Universidad de Bogotá Jorge Tadeo Lozano -Utadeo- se adelantó una caracterización de la población de estudiantes activos y retirados, iniciando con el diseño de dos instrumentos institucionales bajo los nombres de *Hoja de Matrícula* para Nuevos Aspirantes y de actualización de datos para estudiantes antiguos y la *Encuesta de identificación de las causas de deserción estudiantil*. La caracterización de la población de estudiantes activos como identificación de la población también representaba un insumo para analizar el riesgo de deserción, en dos sentidos: el primero operaba en caso de presentarse la no matrícula, porque se tendría información actualizada y detallada posibilitando un análisis integral; el segundo actuaba al tener esta información, en lo posible de forma censal, si se detectaba una característica importante de la población que se retiraba, resultaba imperioso detectar en la población que se retenía cuál era la frecuencia de estudiantes con esta misma característica. En otros términos la información podía ser empleada para delinear perfiles de riesgo y para actuar como contraste en caso de detectar alguna particularidad relevante entre los retirados. Tomando como punto de partida los datos recogidos con los dos instrumentos en 2013-3 y con la Hoja de Matrícula para 2014-1, se realizó un estudio mediante los árboles de decisión con el objeto de jerarquizar las variables explicativas de la variable de interés, el estado *Pago y Matriculado* en 2014-1 en el Sistema Integrado de Información Académica. Para el efecto se examinaron 1.477 datos de ingresos reportados en el periodo 2013-3 y su estado en el 2014-1. Como conclusión en la aplicación del modelo, se obtuvo que una mayor probabilidad de matrícula de un periodo a otro estará adecuadamente distribuida entre quienes tardan un menor tiempo entre la fecha de graduación del colegio y el ingreso a la universidad, aquellos que tengan una experiencia previa en educación superior y que pertenezcan a estratos económicos más altos. Debido al carácter reducido de la muestra, se advierte que el porcentaje de correcta clasificación del modelo es tan solo de 64,39% en la predicción del estado de matrícula. Esta situación es susceptible de mejoría ya que los instrumentos se pueden robustecer, se puede aumentar el tamaño de la muestra de entrenamiento y considerar un mayor número de periodos académicos, todo esto a través de sucesivas aplicaciones.

Descriptor o Palabras Clave: Árboles de Decisión, Caracterización, Contraste, Instrumentos institucionales y Retención.

¹ Profesor Asociado II, Departamento de Ciencias Básicas. Facultad de Ciencias Naturales e Ingeniería. Universidad de Bogotá Jorge Tadeo Lozano. Email: sandra.barragan@utadeo.edu.co

² Profesor Asistente, Departamento de Ciencias Básicas. Facultad de Ciencias Naturales e Ingeniería. Universidad de Bogotá Jorge Tadeo Lozano. Email: leandro.gonzalez@utadeo.edu.co

1. Introducción.

Al inicio de la presente comunicación se ofrece, la descripción y algunos resultados de implementación de los instrumentos: *Hoja de Matrícula* para Nuevos Aspirantes y de actualización de datos para estudiantes antiguos; y *Encuesta de identificación de las causas de deserción estudiantil* como productos de alcance y aplicación institucional en la Universidad de Bogotá Jorge Tadeo Lozano - Utadeo-. En seguida se muestra el estudio de los datos obtenidos mediante la técnica de los árboles de decisión. También se incluyen algunas conclusiones derivadas de las diferentes aproximaciones a los datos con el ánimo de contribuir a la generación de conocimiento alrededor de la caracterización de estudiantes y retirados como otro paso hacia la retención estudiantil.

2. Instrumento para Hoja de Matrícula para Nuevos Aspirantes.

La pretensión de esta encuesta es servir como punto de partida para acopiar de forma sistemática información que permita tener una visión amplia de las características del estudiante, las cuales requieren actualización y adiciones posteriores a medida que avanza su trayectoria académica. Se considera importante caracterizar la población que ingresa a la Utadeo así como a la población activa para identificar los rasgos importantes y para consolidar la información pues en caso de presentarse el evento de la deserción puede

contrastarse las características con las de los estudiantes activos.

El formato base para el instrumento de captura de información proveniente de los Nuevos Aspirantes se diseñó en el Comité Interdisciplinario para la retención estudiantil, teniendo en cuenta las expectativas de las unidades académicas y administrativas que de él toman información. Dichas unidades son: Vicerrectoría Académica, Dirección de Planeación y Finanzas, Dirección de Mercadeo, Dirección de Innovación Educativa y Apoyo Académico, Dirección Académica y Aseguramiento de la Calidad, Coordinación de registro y control académico- Secretaría General, Dirección de Bienestar Universitario, Departamento de Ciencias Básicas, Diseño Gráfico y Dirección de TIC.

En la Hoja de Matrícula se incluyeron las preguntas básicas sobre datos médicos, familiares y académicos, manejo del tiempo libre, hábitos previos al ingreso a la Utadeo y sobre las materias preferidas en el colegio; estas últimas preguntas indagaban sobre la tendencia y la concordancia con la carrera en la que se está tramitando el ingreso con la intención de tener un marco más amplio de la experiencia académica en el bachillerato (Swail, Reed, & Perna, 2003, pág. 87).

2.1. Ficha técnica de la Hoja de Matrícula para el periodo 2013-3.

La Tabla 1 presenta la ficha técnica correspondiente.

Tabla 1. Ficha técnica para la hoja de matrícula para el periodo 2013-3.

Ficha técnica para hoja de matrícula 2013-3	
Nombre de la investigación	Proyecto Tadeísta para la retención estudiantil
Objetivo general	Caracterizar la población de Nuevos Aspirantes a la Utadeo
Tipo de investigación	Encuesta censal
Diseño	Comité interdisciplinario para la retención estudiantil Utadeo.
Análisis	Leandro González Támara. Departamento de Ciencias Básicas
Población objetivo	Nuevos Aspirantes a ingreso a la Utadeo para el periodo 2013-3.
Tamaño de la muestra	892 Nuevos Aspirantes
Técnica de recolección	Diligenciamiento del formulario en el sitio web www.utadeo.edu.co
Fuente	Dirección de Mercadeo-Planeación y Finanzas.
Fecha de recolección	De acuerdo al calendario académico, la inscripción 2013-3, estuvo disponible del 1 de abril hasta el 22 de julio de 2013 y la inscripción de asignaturas para nuevos, reingresos y transferencias externas inició el 14 de mayo hasta la semana anterior a iniciar clases, lo cual se dio el 26 de julio de 2013.

Fuente: Elaboración propia.

2.2. Resultados de la Hoja de Matrícula para Nuevos Aspirantes en el periodo 2013-3.

Se analizaron descriptivamente los 892 datos provenientes de la Hoja de Matrícula destacando los siguientes resultados respecto a la modelación estadística que se hará en la última sección del presente documento:

- Distribución similar por sexo: 461 mujeres y 431 hombres.
- La edad promedio al ingreso es de 20 años (desviación estándar de 3,6 años). Este promedio es ligeramente menor para los aspirantes, 19,5 años.
- Los Nuevos Aspirantes han dejado transcurrir 2,1 años en promedio (desviación estándar de 3,7 años) desde su grado del colegio, 1,6 años para los aspirantes (desviación estándar de 3,3 años). Esto último a pesar de haber obtenido el grado del colegio a los 17,3 años en promedio (desviación estándar de 2 años).
- La distribución del estrato socioeconómico es muy homogénea por programas académicos.

3. Instrumento para Encuesta de identificación de causas de deserción estudiantil en 2013-3.

El formato institucional se diseñó en el Comité Interdisciplinario con el propósito de estimar las circunstancias que influyeron en la decisión de retirarse de la carrera así como su intensidad de la afectación para ofrecer alternativas que propicien un posible reingreso³. Para la formulación de cada pregunta se consideraron: la información sociodemográfica faltante en las bases de datos de la Utadeo y en el módulo específico que hace parte del Sistema Nacional de Información de la Educación Superior - SNIES- (Ministerio de Educación Nacional,

2009, pág. 56), con el nombre de Sistema de Prevención y Análisis de la Deserción en las Instituciones de Educación Superior - SPADIES- y que es alimentado por la información histórica de las Instituciones de Educación Superior. También se acataron la definición de desertor y los determinantes de la deserción estudiantil dados por el Ministerio de Educación Nacional.

El Ministerio de Educación Nacional enuncia los determinantes individuales dados por: edad, género, estado civil, número de hermanos y posición dentro de ellos, problemas de salud, incompatibilidad de horario con las actividades extra clase propuestas para las asignaturas, embarazo, expectativas que no se satisfacen. Los determinantes académicos están relacionados con orientación profesional, tipo de colegio donde culminó el bachillerato, rendimiento académico, resultados en las pruebas de validación, número de asignaturas por periodo académico. Los determinantes institucionales giran alrededor de Becas y apoyos financieros, apoyos académicos, apoyos de bienestar universitario en cuanto a psicología y los determinantes socioeconómicos: Estrato, situación laboral propia y de los padres, personas a cargo, nivel educativo de los padres (Ministerio de Educación Nacional, 2009, pág. 27).

Par brindar soporte a la Encuesta de identificación de causas se elaboró un documento adicional en el que se detallaban las preguntas incluyendo, para cada una de ellas, la intención de su formulación, el determinante al que atañe, si es una variable modificable o no por las actividades de la Utadeo, la base de datos de destino y las acciones asociadas a las posibles respuestas. En el mismo documento, se incluyeron los protocolos de aplicación y procesamiento de los datos obtenidos.

3.1. Ficha técnica para la Encuesta de identificación de causas de deserción estudiantil en el periodo 2013-3.

³ En el diseño de las preguntas se tuvo en cuenta una versión preliminar elaborada en el Proyecto de investigación: La deserción y su relación con la enseñanza de la Ciencias Básicas a cargo del Departamento de Ciencias Básicas, financiado por la Universidad de Bogotá Jorge Tadeo Lozano en convocatoria interna con código 212-05-09. Año 2009.

Tabla 2. Ficha técnica para la Encuesta de identificación de causas de deserción estudiantil para 2013-3.

Ficha técnica para la Encuesta de identificación de causas de deserción estudiantil para 2013-3	
Nombre de la investigación	Proyecto Tadeísta para la retención estudiantil
Objetivo general	Identificar los determinantes de la no matrícula en 2013-3 de la UTadeo.
Tipo de investigación	Encuesta censal
Diseño	Comité interdisciplinario para la retención estudiantil de la Utadeo.
Análisis	Leandro González Támara. Departamento de Ciencias Básicas.
Población objetivo	Estudiantes no matriculados en 2013-3 de los programas académicos ofrecidos por la Universidad, que no radicaron carta de retiro.
Tamaño de la muestra	754 estudiantes no matriculados en 2013-3 que no radicaron carta de retiro en el respectivo programa académico, 129 contestaron la encuesta.
Técnica de recolección	Consulta directa mediante contacto telefónico a través del <i>call center</i> de Información de Programas Académicos –IPA– de la Dirección de Mercadeo de la Utadeo.
Fuente	Dirección de Mercadeo.
Fecha de recolección	Octubre 9 a Noviembre 1 de 2013.

Fuente: Elaboración propia.

3.2. Resultados de la Encuesta de identificación de causas de deserción estudiantil para 2013-3.

La base de datos de no matriculados en 2013-3 contaba con 754 estudiantes que no radicaron carta de retiro de los cuales, por diversos motivos solo se obtuvo información para 129 de ellos (17,1%). Para el procesamiento de la información solo se incluyeron los registros de los estudiantes de los que se tenía información completa para las variables de estudio en la Hoja de Matrícula y en la de identificación de causas de deserción.

Antes de analizar los resultados de la Encuesta de identificación de causas de deserción estudiantil para 2013-3, se calculó el cociente entre Nuevos Aspirantes y No Matriculados en 2013-3 por programa académico. El análisis de esta relación es importante porque refleja parcialmente el aumento o disminución de la masa de estudiantes por programa, independiente del tamaño del mismo. Valores iguales a 1 corresponden a programas donde el número de estudiantes No Matriculados es compensado por los Nuevos Aspirantes. Valores mayores que 1 corresponden a programas donde la masa de estudiantes presentó un aumento y valores menores que 1 corresponde a programas donde la masa presentó una disminución.

Con el procesamiento de la información se concluyó que la no matrícula en 2013-3, se presentó en el momento en que en media los

estudiantes tenían aprobados 62,5 créditos. Derivado de esto se concluyó que si estos estudiantes no matriculados en 2013-3 no renovaban su matrícula en los siguientes dos periodos académicos, en 2014-3 la clasificación promedio con respecto al tiempo⁴ sería deserción temprana.

El estrato no es una variable que identifique a los estudiantes no matriculados en 2013-3 porque mantiene la misma tendencia general.

En cuanto a los motivos de no matrícula: académicos, individuales y económicos, los estudiantes seleccionaron mayoritariamente la categoría marcada como otro, por esta razón se requiere hacer el estudio de los puntos de vista individuales a través de un análisis textual.

4. Un modelo para explicar la no matrícula: Árboles de decisión.

La modelación a través de árboles de decisión permite desarrollar un sistema predictivo que clasifica observaciones enmarcadas en unas reglas de decisión. Esta técnica de modelación estadística, aplicada en el contexto de la deserción estudiantil universitaria, permite clasificar a un estudiante como desertor o no.

⁴ La deserción con respecto al tiempo es clasificada en deserción precoz, en la que el individuo que habiendo sido admitido por la Universidad no se matricula; deserción temprana, en la que se consideran los individuos que abandonan sus estudios en los primeros semestres del programa y en deserción tardía cuando un individuo abandona los estudios en los últimos semestres (Ministerio de Educación Nacional, 2009, págs. 22-23).

Los árboles de decisión son una técnica de estudio de corte transversal, pero permiten ser retroalimentados de manera continua y así generar un sistema de clasificación que cambie con la dinámica misma de la deserción estudiantil (Tan, Steinbach, & Kumar, 2006, págs. 146-148). En cada momento del tiempo un estudiante puede ser considerado como un desertor o no, matriculado o no, en riesgo de desertar o no, o en cualquier otro tipo de clase que sea de interés (Hernández, Ramírez, & Ferri, 2004, pág. 282).

Las reglas de decisión del árbol se basan en la caracterización de cada una de las clases determinadas por la variable de interés que en este caso es si el estado del estudiante en el sistema de la Utadeo es Pago y Matriculado (sí o no) - los estados considerados por la Dirección de TIC son: No pago y matriculado, No pago y no matriculado, No tiene registro en el periodo de observación, Pago y matriculado-. A su vez, el estado Pago y Matriculado se

caracteriza por la información histórica de los estudiantes en los factores que se consideran influyentes.

Por otra parte, la información que no contribuye a la precisión del árbol fue ignorada, lo cual redujo el número de variables.

La Tabla 3 resume algunas de las variables de las que se tiene información para la población matriculada o no en 2014-1 y que fue capturada en 2013-3.

La mencionada información junto con el estado de la matrícula en el primer periodo 2014 pueden ser utilizados para intentar predecir el estado de matrícula de un periodo a otro a través del modelo de árbol de decisión, donde se asume que el estado más deseable es estar matriculado en dicho periodo.

Tabla 3. Listado de variables incluidas en la Hoja de Matrícula.

Listado de variables incluidas en la Hoja de Matrícula	
Fecha de nacimiento	Ha estudiado en otra universidad
Fecha de grado	Semestres en otra universidad
Edad	Consumo de medicamentos con frecuencia
Tipo de ingreso	Discapacidad
Edad del grado del colegio	Proviene de fuera de Bogotá
Sede	Número de aportantes en el hogar
Programa	Ingreso familiar
Sexo	Número de personas en su grupo familiar
Estado civil	Estrato socio económico
País nacimiento	Nivel educativo del padre
Departamento nacimiento	Nivel educativo de la madre
Ciudad de nacimiento	Número de hermanos
Zona residencia	Posición entre los hermanos
Departamento grado del colegio	Personas a cargo
Ciudad grado del colegio	Horas dedicadas a la lectura
Resultado en pruebas ICFES	Deporte que practica
Estudiante trabaja	Años transcurridos desde el grado del colegio hasta la matrícula en la universidad
Parentesco acudiente	

Fuente: Elaboración propia.

A continuación se muestran los resultados de un árbol de Clasificación y Regresión C&R en donde se minimiza la impureza de cada nodo mediante el índice de Gini. Dicho árbol fue obtenido usando el software SPSS⁵. Se omiten los detalles técnicos del proceso para concentrarse en los resultados.

El esquema general del árbol de decisión se muestra en la Fig. 1, en la que se presentan los nodos del 1 al 6.

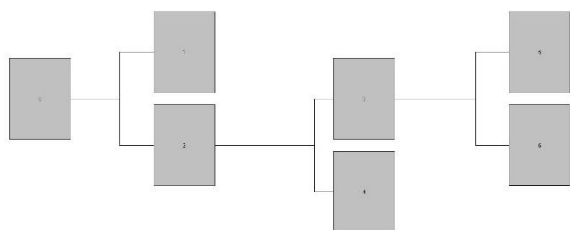


Figura 1. Árbol de decisión; nodos 1 al 6.
Fuente: Elaboración propia.

En la primera etapa, la variable que mejor discrimina los estudiantes que continuaron sus estudios en la Universidad contra los que no lo hacen es el número de años transcurridos desde el año de graduación y su ingreso a la Universidad. Para quienes han ingresado a la universidad dos años y medio o menos después de graduarse del colegio, la tasa de matrícula es de 71,6% mientras que los que lo han hecho en más de dos años y medio, la tasa de matrícula es de 54%. Esta información puede ser usada para predecir el estado de matrícula en un periodo futuro, asumiendo que las condiciones generales de los estudiantes permanecen más o menos constantes. Visto de esta manera, se puede afirmar que, bajo las demás condiciones constantes, la probabilidad de estar matriculado pasa de 0,541 a 0,716 entre quienes han dejado pasar dos y medio años o más contra quienes no. La información detallada se encuentra en la Fig. 2.

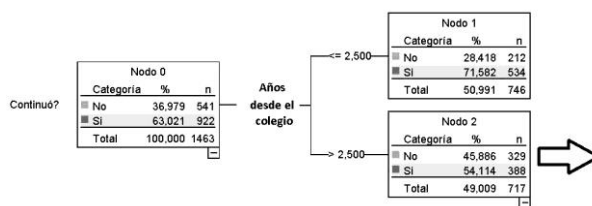


Figura 2. Nodos 0, 1 y 2 del árbol de decisión.
Fuente: Elaboración propia.

Continuando por el nodo 2, es decir estudiantes que han dejado transcurrir dos años o más después del grado del colegio hasta su ingreso a la Utadeo, se concluye que la variable más influyente para discriminar el estado de matrícula es si el estudiante ha estudiado en otra universidad. Si el estudiante ha estudiado en otra universidad, el porcentaje de matrícula es cercano al 60%, mientras que en quienes no han estudiado en otra universidad es de 48%. En la Fig. 3 se muestran los resultados obtenidos.

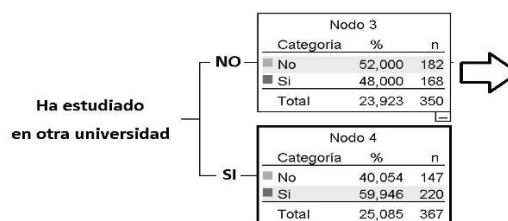


Figura 3. Nodos 3 y 4 del árbol de decisión.
Fuente: Elaboración propia.

Por último y continuando por el nodo 3, la variable que mejor discrimina el estado de matrícula del estudiante es el estrato socioeconómico (estratos 1, 2, 3 y 4 contra 5 y 6). Sin embargo como se aprecia en la Fig. 4, la diferencia entre el porcentaje de matrícula entre estos dos grupos es apenas de 10%.

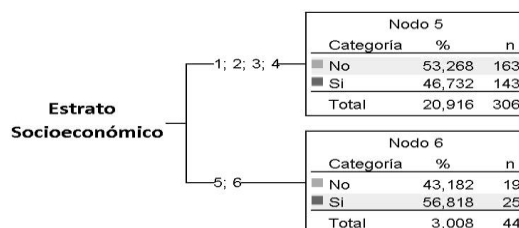


Figura 4. Nodos 5 y 6 del árbol de decisión.
Fuente: Elaboración propia.

Como puede verse, un modelo como el expuesto da una idea de las variables más

⁵ SPSS Inc. (1998). SPSS Base 8.0 for Windows User's Guide. SPSS Inc., Chicago IL.

influyentes en el estado de matrícula de un periodo a otro, dentro de las disponibles en la hoja de matrícula. No obstante la capacidad predictiva del modelo es reducida ya que al evaluar el modelo sobre la misma muestra de entrenamiento, el porcentaje de aciertos es tan solo de 64,39% (ver Fig. 4). Esto puede deberse al desconocimiento de variables de los estudiantes que condicionan con más fuerza el estado de matrícula del estudiante.

Para contar con un mejor modelo se debe aumentar el tamaño de la muestra de entrenamiento, indagar sobre variables influyentes que no hayan sido consideradas (variables financieras, académica, demográficas, etc.) y considerar un mayor número de periodos académicos.

Las reglas de decisión de un modelo como este pueden resumirse en el esquema de la Fig. 5.

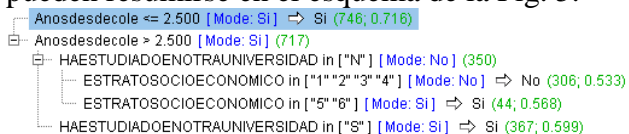


Figura 5. Reglas de decisión para el árbol estudiado.

Fuente: Elaboración propia.

5. Conclusiones.

En estudios de caracterización de estudiantes es importante tener instrumentos institucionales elaborados haciendo coincidir los requerimientos de las unidades académicas y administrativas que tienen relación con el ingreso y la retención.

La caracterización de la población no matriculada debe ser referenciada con la caracterización de la población activa. En caso de encontrar alguna característica relevante en los no matriculados es recomendable verificar entre quienes quedan activos cómo se distribuye dicha característica.

En este estudio de carácter reducido se puede resaltar que las variables más influyentes para explicar el estado de matrícula son:

- Tiempo transcurrido entre el año de graduación del colegio y el de la matrícula en la Universidad.

- Si el estudiante ha realizado estudios en otra universidad.

- El estrato socioeconómico.

Una mayor probabilidad de matrícula de un periodo a otro estará adecuadamente distribuida entre quienes tardan un menor tiempo entre la fecha de graduación del colegio y el ingreso a la Utaeo, aquellos que tengan una experiencia previa en educación superior y que pertenezcan a estratos económicos más altos. Sin embargo estos resultados han de tomarse con cuidado en la predicción del estado de matrícula debido a que el porcentaje de correcta clasificación del modelo es tan solo de 64,39%.

Para contar con un mejor modelo se debe aumentar el tamaño de la muestra de entrenamiento, indagar sobre variables influyentes que no hayan sido consideradas (variables financieras, académica, demográficas, etc.) y considerar un mayor número de periodos académicos.

6. Agradecimientos.

Los autores agradecen a la Universidad de Bogotá Jorge Tadeo Lozano ya que el presente estudio se realizó en el marco del proyecto de investigación: Proyecto Análisis de la retención estudiantil Tadeísta financiado por la Vicerrectoría Académica. También extienden el agradecimiento a la Rectoría por el apoyo otorgado para la participación en el evento CLABES V.

7. Referencias bibliográficas

- Hernández, J., Ramírez, J. M., & Ferri, C. (2004). *Introducción a la minería de datos*. Madrid: Pearson Educación, S.A.
- Ministerio de Educación Nacional. (2009). *Deserción estudiantil en la educación superior colombiana. Metodología de seguimiento, diagnóstico y elementos para su prevención*. Bogotá: Imprenta Nacional de Colombia.
- Swail, W., Reed, K., & Perna, L. (2003). *Retaining minority students in higher education* (Vol. 30). (A. J. Editor, Ed.) Los Angeles: Ashe Eric.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2006). *Introduction to data mining*. Boston: Pearson Education, Inc.